Nucleic-acid-chaperone activity of retroviral nucleocapsid proteins: significance for viral replication

Alan Rein, Louis E. Henderson and Judith G. Levin

Retrovirus particles contain a small, basic protein, the nucleocapsid (NC) protein, that possesses 'nucleic acid chaperone' activity – that is, the NC protein can catalyze the rearrangement of a nucleic acid molecule into the conformation that has the maximal number of base pairs. The molecular mechanism that underlies this effect is not understood. Because the chaperone activity is apparently crucial during the infectious process, NC is a potential target for antiviral therapy.

EXPRESSION OF A single protein species, the Gag polyprotein, is sufficient for assembly of retrovirus particles in cells of higher eukaryotes. If viral RNA containing an intact packaging signal is present in the cell, it will be incorporated into the nascent particle; thus the polyprotein is capable of a highly specific interaction with RNA[1].

Shortly after the virus particle is released from the cell, it undergoes a major structural change termed 'maturation'. Maturation is brought about when the Gag polyprotein is cleaved, by the virus-encoded protease, into a series of products. As indicated in Fig. 1, the cleavage products always include at least three proteins, designated (from N- to C-terminus) the matrix, capsid and nucleocapsid (NC) proteins[1]. The latter is the subject of this review.

Retroviral NC proteins are very small, highly basic proteins: in the case of human immunodeficiency virus type 1 (HIV-1), for example, NC contains only 55 amino acid residues (Fig. 1) and has a pI of 10.0–11.0. NC proteins bind both to ss/dsDNA[2] and to ssRNA *in vitro*, and are associated with the genomic RNA in the interior of the mature retrovirus particle. The NC proteins of all retroviruses (except those of the spumavirus group) contain one or two zinc fingers of the form $CX_2CX_4HX_4C$; according to NMR analysis, these fingers form very tight, rigid loops within the protein. Studies using viral mutants show that these fingers are crucial for packaging of genomic RNA during virus assembly (which occurs before the NC protein is freed from the Gag polyprotein by the viral protease) and that they have other essential functions that have not yet been identified[1,3].

A. Rein is at the Retroviral Genetics Section, ABL-Basic Research Program, NCI-Frederick Cancer Research and Development Center, Frederick, MD 21702, USA; **L. E. Henderson** is at the AIDS Vaccine Program, SAIC Frederick, NCI-Frederick Cancer Research and Development Center, Frederick, MD 21702, USA; and **J. G. Levin** is at the Laboratory of Molecular Genetics, National Institute of Child Health and Human Development, NIH, Bethesda, MD 20892, USA.
Email: rein@mail.ncifcrf.gov

Nucleic-acid-chaperone activity of nucleocapsid proteins

NC proteins have an unusual biochemical activity: they can act as nucleic acid chaperones – that is, they catalyze the folding of nucleic acids into conformations that have the maximal number of base pairs. They appear to do this by lowering the energy barrier for breakage and re-formation of base pairs. Thus, interactions between a nucleic acid molecule and an NC protein result in the transient unpairing of bases within the chain, which makes the bases available for re-pairing in alternative combinations. This activity allows nucleic acid molecules to escape from suboptimal conformations, which would otherwise represent kinetic traps.

This capability of NC proteins has been demonstrated in a wide variety of assay systems: NC proteins accelerate annealing of complementary strands[4–8], facilitate transfer of a nucleic acid strand from one hybrid to a more-stable hybrid[6,8], cause unwinding of tRNA[9] and stimulate release of the products of hammerhead-ribozyme-mediated RNA cleavage[10–12]. Nucleic acid chaperones have been the subject of two recent reviews[13,14]; another protein that has this type of activity is hnRNP A1 (reviewed in Ref. 15).

Work over the past few years has suggested that the nucleic-acid-chaperone activity of NC protein comes into play at at least three points during the retroviral life cycle and is of crucial importance for viral replication. In this review, we discuss this activity and its biological significance. Although much of the work presented has been done using the HIV-1 NC protein, the conclusions drawn presumably apply to other retroviruses as well.

Maturation of dimeric RNAs within retrovirus particles

In all retrovirus particles, the genomic RNA is actually present as a dimer. The structure of this dimer is not well understood; however, the dimer is known to be composed of two identical, positive-strand, genomic RNA molecules that are held together by a limited number of base pairs[1]. Studies on several different retroviruses have shown that the dimer undergoes a conformational change within the first hour(s) after the virus is released from the cell; this change renders it more compact (it migrates more rapidly in both sedimentation and electrophoresis analyses) and more thermostable (the temperature required to dissociate the dimer increases)[16–18]. This conformational change is termed maturation of the dimer. It seems likely that, initially, the genomic RNA is incorporated into the nascent particle in the form of a dimer and that subsequently it is converted to a new, more-stable dimeric state – presumably one in which there is a higher number of base pairs.

Studies on mutant viruses were employed in an attempt to elucidate the molecular mechanisms that underlie the maturation event[17,18]. These experiments showed that maturation of the dimer is dependent upon the presence of an active virus-encoded protease within the particle. This finding suggested the existence of a three-step pathway: (1) an immature dimer is packaged into a virus particle; (2) the viral protease, activated when the virus is released from the cell,

 0968–0004/98/$19.00 PII: S0968-0004(98)01256-0

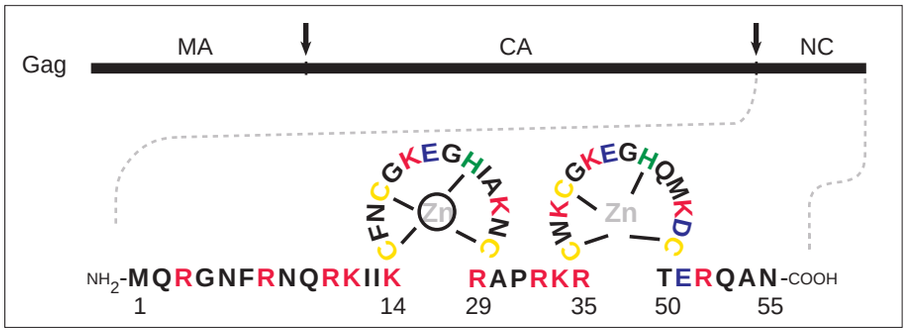

Figure 1

Origin and structure of the nucleocapsid (NC) protein. During maturation of a retrovirus particle, the Gag polyprotein is cleaved (indicated by arrows) into a series of products. The products always include the matrix (MA), capsid (CA) and NC proteins, although one or more non-conserved products are often present as well. The sequence of the MN isolate of HIV-1 NC is shown. Basic residues are shown in red; acidic residues are shown in blue; zinc-coordinating residues are shown in yellow or green.

cleaves the Gag polyprotein into a series of products; and (3) one of these cleavage products interacts with the immature dimeric RNA, enabling the RNA to adopt a more-stable dimeric form. Because NC proteins associate with the viral RNA within the mature retrovirus particle and possess nucleic-acid-chaperone activity *in vitro*, it was natural to postulate that the NC protein is the cleavage product responsible for maturation of the dimeric RNA.

This hypothesis was strongly supported by experiments using recombinant HIV-1 NC protein *in vitro*. These experiments exploited the fact that RNA transcripts that are a few hundred bases in length, and contain sequences from the 5′ end of retroviral genomes, can dimerize spontaneously when incubated in buffers that have a high ionic strength and contain Mg^{2+} (Ref. 4). Depending upon the incubation conditions used, 335-nucleotide transcripts of Harvey-sarcoma-virus RNA form one of two alternative types of dimer, which differ in their thermostability. This observation made it possible to show that incubation of the less-stable form of dimer with recombinant NC protein caused conversion of the dimer to the more stable form[19]. Analogous observations were also made in a study that used HIV-1, rather than Harvey sarcoma virus, transcripts[20]. The conformational change induced in these transcripts by the NC protein appears to mimic, in a simple, defined system *in vitro*, the maturation of dimeric RNA that takes place within the retrovirus particle.

The effect of nucleocapsid protein upon the efficiency of reverse transcription

When DNA is generated from an RNA template *in vitro* by reverse transcriptase (RT), the DNA product is frequently incomplete – that is, RT tends to 'pause' or 'stall' at specific sites on the template RNA[21–26]. These sites are often, though not always, stem-loop structures in the RNA. The presence of recombinant HIV-1 NC protein greatly reduces pausing by RT and increases the efficiency of synthesis of the full-length DNA product[26–28]. This effect almost certainly reflects the nucleic-acid-chaperone activity of the NC protein. Because the protein can break base pairs, it appears to be able to eliminate transiently the secondary structures in the template RNA that obstruct the polymerization process (Fig. 2). Although this effect of NC protein has only been demonstrated *in vitro*, it is probably quite significant during the viral DNA synthesis that occurs upon infection of a new host cell by a retrovirus particle.

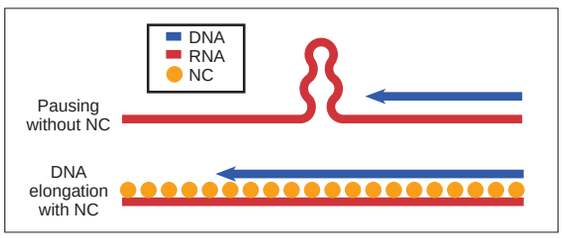

Figure 2

Effect of nucleocapsid (NC) protein on DNA elongation during reverse transcription. Addition of NC to reverse transcription reactions *in vitro* reduces the pausing of reverse transcriptase, presumably by transiently destabilizing regions of secondary structure in the template RNA.

The effect of nucleocapsid protein on strand transfer during reverse transcription

Synthesis of the DNA copy of the viral genome is initiated from a cellular tRNA molecule. The 3′ 18 bases of this tRNA are complementary to a region of the viral RNA termed the primer-binding site (PBS); the tRNA is annealed to the PBS during (or before) assembly of the virus particle. Somewhat surprisingly, the PBS is only a short distance (~100–200 bases) from the 5′ end of the genomic-RNA template. The template is terminally redundant; the sequences that are repeated at each end of the viral RNA are called the R region. The first product of reverse transcription is a DNA copy of the ~100–200 bases that lie between the PBS and the 5′ terminus of the viral RNA (see Fig. 3). As the RNA template is copied, it is digested by the RNase H activity of reverse transcriptase. The DNA fragment produced, termed the '(−) strand strong-stop DNA', must then be annealed to the R region at the 3′ end of the genomic RNA before reverse transcription can continue. This annealing event is termed 'strand transfer'[1].

In the case of HIV-1, attempts to reconstruct the reverse-transcription process using purified components have shown that strand transfer is extremely inefficient *in vitro* if the 5′ end of the template RNA is a faithful replica of the 5′ end of the viral genome. Analysis of the products revealed that this inefficiency is due to a competing synthetic pathway: the 3′ end of the (−) strand strong-stop DNA folds into a stable secondary structure composed of sequences that are complementary to the TAR-stem loop that is present at the 5′ end of the HIV-1 genome. This secondary structure enables the DNA to fold back on itself and self-prime DNA synthesis[29,30]; this whole process leads to the addition of (+) strand sequences to the (−) strand strong-stop DNA (Fig. 3). Formation of these extension products is incompatible with successful strand transfer[30] and with synthesis of a faithful DNA copy of the viral genome.

The presence of NC protein profoundly increases the efficiency of strand transfer by HIV-1 reverse transcriptase: the level of transfer rises from ~3% (in the absence of NC protein) to as much as 65% [in the presence of high levels of NC protein (see Fig. 3)][30–34]. This effect of the NC protein, like its effect on the efficiency of strand elongation, can be ascribed to its nucleic-acid-chaperone activity. Thus, the hybrid between the R sequence at the 3′ end of genomic RNA and its complement (r) in the (−) strand strong-stop DNA contains more base pairs than does the

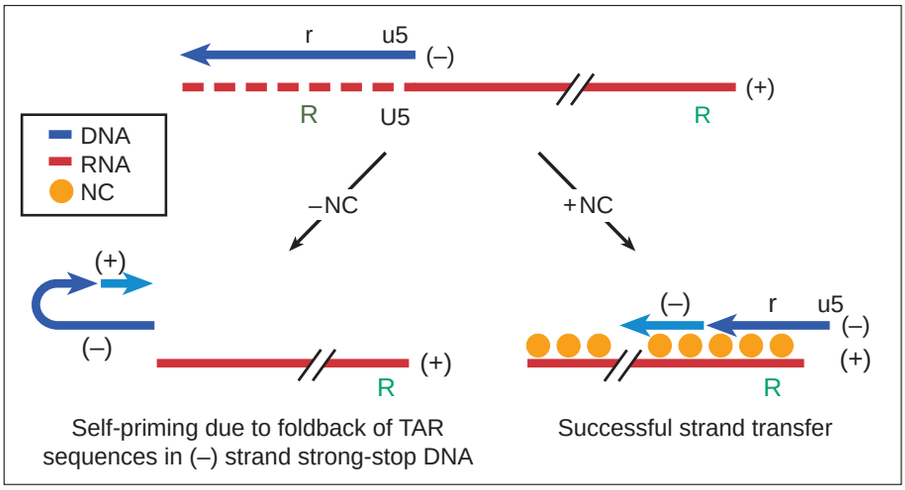

Figure 3

Effect of nucleocapsid (NC) protein on strand transfer during reverse transcription. Synthesis of a complete (−) strand DNA copy of the viral genomic RNA depends on the annealing of 'r' sequences in the (−) strand strong-stop DNA (the initial product of reverse transcription) to R sequences at the 3′ end of the template RNA. However, in a competing reaction, self-complementary sequences (the complement of the TAR stem-loop) cause the (−) strand strong-stop DNA to fold back on itself and prime continued synthesis, which results in the extension of the (−) strand DNA with (+) strand sequences. Addition of NC to reverse transcription reactions *in vitro* prevents this self-priming reaction and profoundly increases the level of synthesis of full-length (−) strand DNA.

stem-loop that is formed by the (−) strand TAR sequence in the (−) strand strong-stop DNA. We can assume that the NC protein destabilizes the (−) strand TAR stem-loop in the DNA, as well as the TAR sequences in the 3′ end of the template RNA; this makes their constituent bases available for formation of the more stable r–R hybrid that is required for successful strand transfer[7,30]. These *in vitro* results, like those on strand elongation, strongly suggest that NC is a crucial accessory protein for reverse transcription *in vivo*.

RNA–RNA interactions before/during retrovirus assembly: a possible nucleic-acid-chaperone activity of the Gag polyprotein

Analysis of virus particles that lack the viral protease shows that at least two RNA–RNA interactions occur before or during assembly of the retrovirus particle (i.e. before cleavage of the Gag polyprotein and release of the NC protein)[17,18,35–38]: the initial association of two genomic RNA molecules to form the immature dimer; and the annealing of a cellular tRNA molecule to the PBS. Because the latter entails the disruption of the normal secondary and tertiary structure of the tRNA, it would seem to require an as-yet-unidentified nucleic-acid-chaperone cofactor. This putative cofactor appears to have some relationship to the NC protein: an early study showed that a mutation within the NC coding region prevented placement of the tRNA on the PBS (Ref. 39). However, even though the NC protein is capable of catalyzing both dimerization and tRNA annealing *in vitro*[4], the NC protein per se is not required for these events *in vivo* – as noted above[17,18,35–38], particles that lack an active protease still contain immature dimeric RNA molecules with annealed tRNAs. An attractive possibility is that the Gag polyprotein itself, in which the NC peptide exists as a domain, is responsible for these effects. This seems particularly plausible, because several intermediate cleavage products of Gag, which contain NC along with other, adjacent domains, possess chaperone activity[5,40–42].

Molecular mechanisms that underlie the nucleic-acid -chaperone activity of nucleocapsid protein

The ability of the NC protein and other nucleic acid chaperones to destabilize base pairs transiently, and thus catalyze the rearrangement of nucleic acids into optimally base-paired conformations, has not yet been explained in molecular terms. However, because of its small size (the entire 55-amino-acid sequence of the HIV-1 NC protein is shown in Fig. 1), the NC protein would appear to offer an ideal system for mechanistic investigations.

One approach to understanding the mechanism of nucleic-acid-chaperone action has been that of analyzing the stoichiometry of the reaction. Dose-response studies have shown consistently that a threshhold concentration of the protein is required for demonstrable nucleic-acid-chaperone effects[7,9,19,26,30]. This minimum concentration is generally in the range of 1 protein molecule per ~7 nucleotides. Because binding measurements have shown that this is approximately the ratio at which a nucleic acid becomes saturated with NC protein[5,43,44], it appears that there is little chaperone effect unless the nucleic acids are almost completely covered by the protein. This in turn raises the possibility that there is no turnover in the catalysis of conformational changes by NC protein. (The need for saturating amounts of NC protein might also indicate that interactions between NC-protein molecules bound to nucleic acid play a role in the chaperone effect, as suggested by Tanchou *et al.*[28])

The mechanism behind nucleic-acid-chaperone activity has also been investigated by testing the activity of mutant NC proteins, and a number of different assays have been used. These studies have shown that the highly basic character of the NC protein is essential for its activity. By contrast, drastic alterations in, or even elimination of, the zinc fingers (e.g. replacing all six cysteine residues in the two HIV-1 NC zinc fingers with alanine residues) generally cause a somewhat-limited reduction in the chaperone activity of the protein[12,19,26,41].

The ability of the NC protein to catalyze conformational rearrangements in nucleic acids is, as far as is known, independent of the sequence of the nucleic acid, and many assays for chaperone activity use nonviral nucleic acid sequences. Nevertheless, the binding of the NC protein to nucleic acids – as measured by surface plasmon resonance, filter binding, footprinting or electrophoretic-mobility-shift assays – exhibits strong sequence preferences[3,45–48]. These preferences could reflect the specificity with which the parental Gag polyprotein encapsidates genomic RNA during virus assembly, but they might also relate to the functions of the NC protein itself as a nucleic acid chaperone. It would be intriguing to know whether or not the sequence specificity of interactions between the NC protein and nucleic acids affects the ability of NC protein to destabilize base pairs transiently – as required for its nucleic-acid-chaperone activity.

Very recently, the three-dimensional

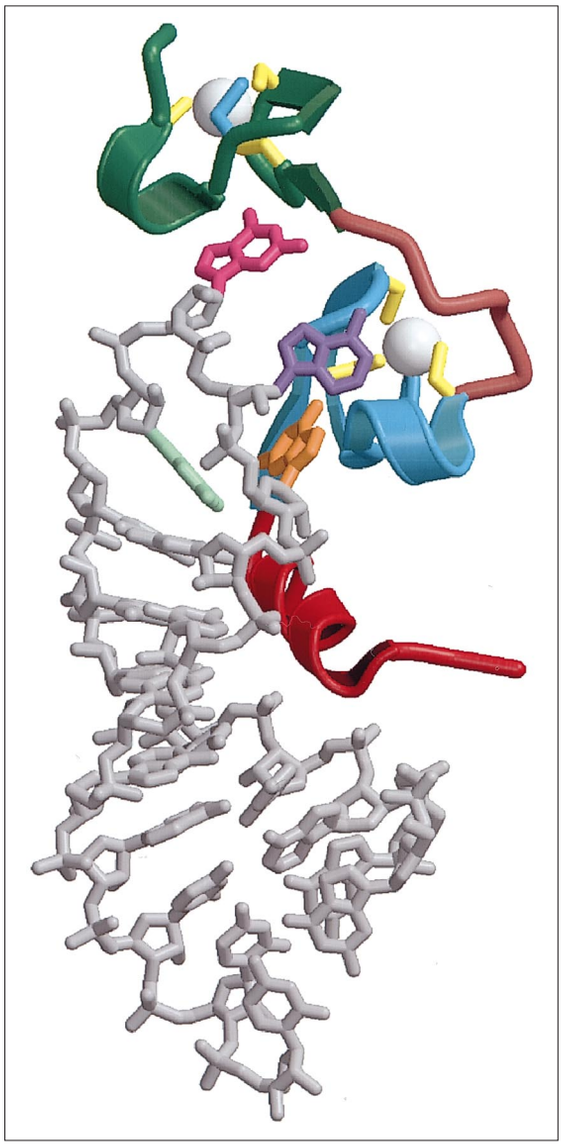

Figure 4

Three-dimensional NMR structure of the nucleocapsid (NC) protein bound to an RNA stem-loop. The N-terminal 12 residues of NC (shown in red) lie in the major groove of the RNA stem (shown in gray), while the two zinc fingers (shown in blue and green) engage in specific interactions with several of the purine bases (shown in green, pink, purple and orange) in the loop. The zinc ions are shown in white. Coordinating cysteine residues are shown in yellow; coordinating histidine residues are shown in blue. The NC protein shown is from the NL4-3 isolate of HIV-1. Figure courtesy of Michael F. Summers and Roberto N. De Guzman.

structure of a complex between an HIV-1 NC protein and an RNA stem-loop was obtained by NMR[49] (Fig. 4). The stem-loop is part of the signal (in the viral RNA) that directs incorporation of the RNA into the assembling virus particle, and it binds to NC protein with relatively high affinity ($K_d \approx 100$ nM). In the complex, the zinc fingers (see Fig. 1) both engage in highly specific interactions with bases in the RNA loop. By contrast, residues 3–11 (at the N-terminus of the protein), which include four conserved basic residues (Lys3, Arg7, Arg10 and Lys11), form a 3_{10}-helix that penetrates the major groove of the RNA stem [note that the NC protein used in the NMR study (Fig. 4) is from an HIV-1 isolate that is different from that shown in Fig. 1; there are minor differences in the amino acid sequences of the two proteins]. These basic residues all participate in nonspecific, electrostatic interactions with the phosphodiester groups of the stem. In addition, Asn5 interacts with a G base and a C base that lie near the top of the stem in this structure.

Strikingly, the NC protein appears to have two distinct domains: one (the zinc fingers) engages in highly specific interactions with bases; the other (the N-terminal basic region) binds to an RNA helix, largely through electrostatic interactions with the backbone. However, the protein could well adopt other conformations in complexes that involve other nucleic acid structures[50]; as noted above, chaperone activity is apparently independent of the nucleic acid sequence in the complex. The data reviewed above are generally consistent with the hypothesis that the basic, N-terminal domain is responsible, at least in part, for the nucleic-acid-chaperone activity of the protein, although the mechanism of this activity is still an unsolved problem.

Implications

The fundamental conclusion from the work discussed here is that retroviruses possess a protein, the NC protein, that performs crucial catalytic functions during viral maturation and infection. In this respect, the NC protein resembles the three classical retroviral enzymes – protease, RT and integrase – and, like these enzymes, the NC protein can be considered to be a potential target for antiviral therapy.

Such therapies might be of two types: agents that interfere with NC-catalyzed events without interacting with the NC protein itself; and agents that attack the NC protein directly. One example of the first class is actinomycin D, which inhibits the annealing step in HIV-1 (−) strand-transfer *in vitro*, presumably as a result of an interaction between actinomycin D and the (−) strand strong-stop DNA. NC protein was unable to overcome this effect of actinomycin D (Ref. 51).

Several other compounds that appear to inactivate retroviruses by interacting with the NC protein have been identified[52,53]. These compounds are hydrophobic, mild oxidizing agents; their hydrophobic character enables them to cross the lipid bilayer that coats the virus particle, while their oxidizing activity induces disulfide crosslinking between Cys residues in the NC-protein zinc fingers within the viral core. Despite their lack of chemical specificity, these agents are relatively nontoxic to the host, probably because mild oxidizing agents are rendered innocuous by the reducing environment within the host cell.

The absolute conservation of the $CX_2CX_4HX_4C$ motif in retroviral NC proteins, and the fact that all known mutations in this motif are lethal for the virus, raises the possibility that the virus will be unable to generate mutants that are resistant to these hydrophobic, mild oxidizing agents. Further, the conservation of the motif among unrelated retroviruses means that murine models of retrovirus disease can be used for initial evaluation of potential antiviral agents. Such a rapid, inexpensive screen is obviously essential, given the extraordinarily large number of chemicals that fit the general description of 'hydrophobic, mild oxidizing agents'. In fact, one such agent has been shown to delay the onset of retrovirus-induced disease in mice[54]. It would seem important to extend the testing of these oxidizing agents to better models of HIV-induced disease or into the clinic. However, the development of other approaches for inactivating retroviruses by targeting the NC protein is an extremely worthwhile goal; further understanding of the molecular mechanisms that underlie nucleic-acid-chaperone activity should make an invaluable contribution to this effort.

Acknowledgements

We thank Jianhui Guo and Robert Gorelick for helpful comments on the manuscript. A. R. is supported in part by the National Cancer Institute, DHHS, under contract with ABL. J. G. L. is supported in part by the NIH Intramural AIDS Targeted Antiviral Program.

References

1 Coffin, J. M., Hughes, S. H. and Varmus, H. E. (1997) *Retroviruses*, Cold Spring Harbor Laboratory Press
2 Lapadat-Tapolsky, M. *et al.* (1993) *Nucleic Acids Res.* 21, 831–839
3 Berkowitz, R., Fisher, J. and Goff, S. P. (1996) *Curr. Top. Microbiol. Immunol.* 214, 177–218
4 Darlix, J. L., Lapadat-Tapolsky, M., de Rocquigny, H. and Roques, B. P. (1995) *J. Mol. Biol.* 254, 523–537
5 Dib-Hajj, F., Khan, R. and Giedroc, D. P. (1993) *Protein Sci.* 2, 231–243
6 Tsuchihashi, Z. and Brown, P. O. (1994) *J. Virol.* 68, 5863–5870
7 You, J. C. and McHenry, C. S. (1994) *J. Biol. Chem.* 269, 31491–31495
8 Lapadat-Tapolsky, M., Pernelle, C., Borie, C. and Darlix, J-L. (1995) *Nucleic Acids Res.* 23, 2434–2441
9 Khan, R. and Giedroc, D. P. (1992) *J. Biol. Chem.* 267, 6689–6695
10 Bertrand, E. L. and Rossi, J. J. (1994) *EMBO J.* 13, 2904–2912
11 Herschlag, D., Khosla, M., Tsuchihashi, Z. and Karpel, R. L. (1994) *EMBO J.* 13, 2913–2924
12 Muller, G. *et al.* (1994) *J. Mol. Biol.* 242, 422–429
13 Herschlag, D. (1995) *J. Biol. Chem.* 270, 20871–20874
14 Weeks, K. M. (1997) *Curr. Opin. Struct. Biol.* 7, 336–342
15 Dreyfuss, G., Matunis, M. J., Pinol-Roma, S. and Burd, C. G. (1993) *Annu. Rev. Biochem.* 62, 289–321
16 Stoltzfus, C. M. and Snyder, P. N. (1975) *J. Virol.* 16, 1161–1170
17 Fu, W. and Rein, A. (1993) *J. Virol.* 67, 5443–5449
18 Fu, W., Gorelick, R. J. and Rein, A. (1994) *J. Virol.* 68, 5013–5018
19 Feng, Y-X. *et al.* (1996) *Proc. Natl. Acad. Sci. U. S. A.* 93, 7577–7581
20 Muriaux, D., De Rocquigny, H., Roques, B. P. and Paoletti, J. (1996) *J. Biol. Chem.* 271, 33686–33692
21 Messer, L. I. *et al.* (1985) *Virology* 146, 146–152
22 Dudding, L. R., Nkabinde, N. C. and Mizrahi, V. (1991) *Biochemistry* 30, 10498–10506
23 DeStefano, J. J. *et al.* (1992) *Biochim. Biophys. Acta* 1131, 270–280
24 Klarmann, G. J., Schauber, C. A. and Preston, B. D. (1993) *J. Biol. Chem.* 268, 9793–9802
25 Arts, E. J., Li, Z. and Wainberg, M. A. (1995) *J. Biomed. Sci.* 2, 314–321
26 Wu, W. *et al.* (1996) *J. Virol.* 70, 7132–7142
27 Ji, X., Klarmann, G. J. and Preston, B. D. (1996) *Biochemistry* 35, 132–143
28 Tanchou, V., Gabus, C., Rogemond, V. and Darlix, J-L. (1995) *J. Mol. Biol.* 252, 563–571
29 Lapadat-Tapolsky, M., Gabus, C., Rau, M. and Darlix, J-L. (1997) *J. Mol. Biol.* 268, 250–260
30 Guo, J. *et al.* (1997) *J. Virol.* 71, 5178–5188
31 Darlix, J-L. *et al.* (1993) *C. R. Acad. Sci. (Ser. III)* 316, 763–771
32 Allain, B., Lapadat-Tapolsky, M., Berlioz, C. and Darlix, J-L. (1994) *EMBO J.* 13, 973–981
33 Peliska, J. A., Balasubramanian, S., Giedroc, D. P. and Benkovic, S. J. (1994) *Biochemistry* 33, 13817–13823
34 Kim, J. K. *et al.* (1997) *J. Biol. Chem.* 272, 16769–16777
35 Crawford, S. and Goff, S. P. (1985) *J. Virol.* 53, 899–907
36 Stewart, L., Schatz, G. and Vogt, V. M. (1990) *J. Virol.* 64, 5076–5092
37 Fu, W. *et al.* (1997) *J. Virol.* 71, 6940–6946
38 Huang, Y. *et al.* (1997) *J. Virol.* 71, 726–728
39 Prats, A. C. *et al.* (1988) *EMBO J.* 7, 1777–1783
40 Darlix, J-L. *et al.* (1990) *J. Mol. Biol.* 216, 689–699
41 De Rocquigny, H. *et al.* (1992) *Proc. Natl. Acad. Sci. U. S. A.* 89, 6472–6476
42 Li, X. *et al.* (1996) *J. Virol.* 70, 4996–5004
43 Karpel, R. L., Henderson, L. E. and Oroszlan, S. (1987) *J. Biol. Chem.* 262, 4961–4967
44 You, J. C. and McHenry, C. S. (1993) *J. Biol. Chem.* 268, 16519–16527
45 Dannull, J., Surovoy, A., Jung, G. and Moelling, K. (1994) *EMBO J.* 13, 1525–1533
46 Allen, P. *et al.* (1996) *Virology* 225, 306–315
47 Berglund, J. A., Charpentier, B. and Rosbash, M. (1997) *Nucleic Acids Res.* 25, 1042–1049
48 Fisher, R. J. *et al.* (1998) *J. Virol.* 72, 1902–1909
49 De Guzman, R. N. *et al.* (1998) *Science* 279, 384–388
50 Frankel, A. D. and Smith, C. A. (1998) *Cell* 92, 149–151
51 Guo, J. *et al.* (1998) *J. Virol.* 72, 6716–6724
52 Rice, W. G. *et al.* (1995) *Science* 270, 1194–1197
53 Rein, A. *et al.* (1996) *J. Virol.* 70, 4966–4972
54 Ott, D. E. *et al.* (1998) *Virology* 243, 283–292

Gathering STYX: phosphatase-like form predicts functions for unique protein-interaction domains

Matthew J. Wishart and Jack E. Dixon

The effects of tyrosine phosphorylation are manifested and regulated through protein domains that bind to specific phosphotyrosine motifs. STYX is a unique modular domain found within proteins implicated in mediating the effects of tyrosine phosphorylation *in vivo*. Individual STYX domains are not catalytically active; however, they resemble protein tyrosine phosphatase (PTP) domains and, like PTPs, contain core sequences that recognize phosphorylated substrates. Thus, the STYX domain adds to the repertoire of modular domains that can mediate intracellular signaling in response to protein phosphorylation.

PROTEIN PHOSPHORYLATION ON tyrosine residues is a critical step in the regulation of processes such as metabolism, proliferation and differentiation. Its effects arise both directly, through alterations in protein structure or enzyme activity, and indirectly, through facilitation of protein–protein interactions. The phosphorylation state of a protein is regulated by the activity of protein tyrosine phosphatases (PTPs) and protein tyrosine kinases (PTKs) that hydrolyze and attach phosphate groups, respectively (Fig. 1), to amino acid residues in specific short sequence motifs[1].

In addition to being substrates for PTPs, phosphotyrosine (pY) motifs serve as binding sites for non-catalytic protein domains that are structurally unrelated to phosphatases. Src-homology 2 (SH2) and phosphotyrosine-interaction/binding (PI/PTB) domains mediate phosphorylation-dependent association with specific pY motifs in target proteins (Fig. 1)[2,3]. In some cases, these domains can compete with PTPs for binding to a site and thereby prevent its dephosphorylation[4]. Thus, the manner in which tyrosine phosphorylation effects are manifested and regulated reflects an interplay between proteins with non-catalytic binding domains and proteins with hydrolytic binding domains. Identification of new pY-directed protein domains would increase the range of potential functions and regulatory mechanisms initiated by individual phosphorylation events.

Here, we describe a unique collection of modular protein sequences – the STYX domains (also known as STYX-phosphatase domains). Although related to the PTP superfamily, STYX domains lack phosphatase activity because of naturally occurring substitutions of catalytically essential residues. Rather than simply being non-functional homologs of active phosphatases, STYX domains

M. J. Wishart is at the Dept of Physiology, University of Michigan, Ann Arbor, MI 48109-0606, USA; and **J. E. Dixon** is at the Dept of Biological Chemistry, University of MI, Ann Arbor, MI 48109-0606, USA.

 0968–0004/98/$19.00 PII: S0968-0004(98)01241-9